2. Materials and Methods

2.1. Daily Confirmed Dengue Cases

We derived data on confirmed Dengue cases from the Directorate General of Health Services (DGHS)'s website.

2.2. Meteorological Factors

We used NASA’s Prediction of Worldwide Energy Resources webpage (NASA, 2022) on a daily interval to collate meteorological variables including Rainfall (mm), Relative humidity [RH, (%)], Temperatures (°C), Surface pressure (kPa), Dew point (°C), and wind velocity (m/s) at 10 m height (Maximum Wind Speed) available from https://power.larc.nasa.gov/data-access-viewer (Supplementary data SD2).

2.3. Time Series Models

The sampled variables were used to choose Six promising time series models namely SES, ARIMA, SARIMA, ARIMAX, GA and Prophet employed to examine the trend in dengue cases. We forecasted the number of new dengue cases for 30 subsequent days while using the SES model as a benchmark to assess the prediction accuracy of other models.

2.3.1. SES Model (Simple Exponential Smoothing):

The SES model is the most used model for analyzing time series datasets (de Livera et al., 2011). The SES is an easy-to-use tool that treats data as varying around a constant mean (Tseng & Shih, 2020). In this technique, more recent observations get higher weightage. That is exponential weights assign to the observations to make predictions . The simplest of exponential smoothing is simple exponential smoothing which is appropriate when no trend or seasonal pattern exists (https://www.researchgate.net/profile/Shyam-Perera/publication/338040747\_COMBINING\_FORECASTS\_OF\_ARIMA\_AND\_EXPONENTIAL\_SMOOTHING\_MODELS/links/5f1550394585151299aad015/COMBINING-FORECASTS-OF-ARIMA-AND-EXPONENTIAL-SMOOTHING-MODELS.pdf). It is a univariate time-series forecasting approach in which there is no evidence of trend or seasonality in the data. The SES model in this study is performed using the R package “fpp2”(Chaurasia & Pal, 2020).

The equation of the SES model can be expressed as:

F\_t=F\_(t-1)+α(A\_(t-1)-F\_(t-1)) (1)

Where〖 A〗\_t is the actual value of the series at time t, F\_t is the forecast value of the series at time t, and α is a weighting parameter that takes a value between 0 and 1.

2.3.2. ARIMA Model (Auto-Regressive Integrated Moving Average):

ARIMA is one of the most widely applied time series analysis techniques which captures seasonal, trend and other cyclic patterns in the data to forecast the future values of the series. Model identification, parameter estimation and model diagnostic checking are the three stages in ARIMA modelling. After seasonality and stationarity have been identified, autoregressive and moving average components are identified using autocorrelation (ACF) and partial autocorrelation functions. In model diagnostic checking residuals should follow a white noise which is drawn from a constant mean and variance. If the assumptions are violated, then another model needs to be investigated. To produce stationary time series, the general linear stochastic model incorporates autoregressive, moving-average models, and differencing (Adhikari & Agrawal, 2013). In a conventional autoregressive model, the future values of variables of interest are predicted using a linear combination of the past values. Like a regression model, the moving-average model incorporates the errors from past projections. The ARIMA produces accurate results in the absence of seasonality in the data (Tariq et al., 2021). The R package “forecast” is utilized in this study to run the ARIMA model expressed as (de Livera et al., 2011):

y\_t=c+〖φ\_1 y〗\_(t-1 )+⋯+〖φ\_p y〗\_(t-p)+〖θ\_1 e〗\_(t-1 )+⋯+〖θ\_q e〗\_(t-q)+e\_t(2)

where yt is the difference at degree d of the original series of time series, φ\_1-φ\_p are autoregressive model parameters, θ\_1-θ\_q represents moving-average model parameters, and e\_t is white noise.

2.3.3. Prophet Model (Automatic Forecasting time-series model):

The Prophet model fits the data set relatively quickly and is utilized for observations with irregularities, a model for additive regression. The Prophet model’s components are a logistic gain curve trend, which selects the data’s growing points to identify variances in trends. The Fourier series can be used to simulate an annual seasonal component whereas dummy variables can represent a seasonal weekly element. The model functions effectively with historical data from various seasons and time series with significant seasonal fluctuation while accounting for missing data and outliers. As a result, time series generally benefit from this without losing data values or transferring in a trend or outlier. The Prophet model was introduced by Taylor and Letham (2017). The Prophet model with extra regressor (Taylor, S.J., Letham, B. 2017. Forecasting at scale. The American Statistician, 72 (1), 37- 45. https://doi.org/10.1080/00031305.2017.1380080. Taylor, S.J., Letham, B. 2019. Prophet: Automatic Forecasting Procedure, R package version 0.5, available at https://cran.r-project.org/web/packages/Prophet/index.html (last access: 1April 2019) has three components: trend, seasonality and regressor with an error term. The R package “Prophet” is used to analyze time-series data using the Prophet model to track cases and fatalities of dengue expressed as (Hasan et al., 2021):

y\_t=g\_t+s\_t+h\_t+ ∈\_t(3)

Equation 3 is used for Prophet analysis whereg\_t, s\_t, and h\_t are model factors and ∈\_t is used for non-periodic changes.

2.3.4. Autoregressive Integrated Moving Average with Explanatory Variables (ARIMAX)

Furthermore, it can incorporate external factors, such as climatic variables, which can improve the fitting and prediction accuracy; under these circumstances, the model as an extension of the ARIMA model is named the ARIMAX model (https://www.cambridge.org/core/journals/epidemiology-and-infection/article/imported-cases-and-minimum-temperature-drive-dengue-transmission-in-guangzhou-china-evidence-from-arimax-model/26D6CE21F73579E5DF00EF720845EC62). The ARIMA model accepts a direct relationship between the time-series values attempts to exploit these straight conditions in perceptions, and arranges to extricate nearby designs, removing high-frequency commotion. In this model, the explanatory information variable (X) is added, which is called ARIMAX (p, d, q) for accurate interpretation (Adhikari & Agrawal, 2013). Wangdi et al. [K. Wangdi, P. Singhasivanon, T. Silawan, S. Lawpoolsri, N. J. White, and J. Kaewkungwal, “Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan,” Malaria Journal, vol. 9, article 251, 2010.] adapted ARIMAX model to determine predictors of malaria for the subsequent month. And the test showed that prediction accuracy has been greatly improved. Chadsuthi et al. [S. Chadsuthi, C. Modchang, Y. Lenbury, S. Iamsirithaworn, and W. Triampo, “Modeling seasonal leptospirosis transmission and its association with rainfall and temperature in Thailand using time-series and ARIMAX analyses,” Asian Pacific Journal of Tropical Medicine, vol. 5, no. 7, pp. 539–546, 2012.] studied seasonal leptospirosis transmission and the association with rainfall and temperature by using ARIMAX model showing that factoring in rainfall (with an 8-month lag) yields the best model for the northern region.

SARIMA

The SARIMA model takes both overall trends and seasonal changes into account, which is widely used in modeling time series [13–15]. We established and selected the best SARIMA model (p, d, q) × (P, D, Q) according to the steps introduced by Box and Jenkins [Ziegler, T.; Mamahit, A.; Cox, N.J. 65 years of influenza surveillance by a World Health Organization-coordinated global network. Influenza Other Resp. Viruses 2018, 12, 558–565.] (Figure 1). Autoregressive lags, moving average lags, seasonal autoregressive lags, and seasonal moving average lags are indicated by p, q, P, and Q, respectively.

GAM

Generalized additive models (GAMs) (Hastie TJ, Tibshirani RJ. Generalized additive models. New York, NY: Chapman and Hall, Inc, 1990.) have been effectively applied in a variety of research areas. In time-series studies of air pollution and mortality, GAM has been the most widely applied method, because it allows for nonparametric adjustments for nonlinear confounding effects of seasonality, trends, and weather variables (Katsouyanni K, Toulomi G, Samoli E, et al. Confounding and effect modification in the short-term effects of ambient particles on total mortality: results from 29 European cities within the APHEA2 project. Epidemiology 2001;12:521–31.). It is a more flexible approach than fully parametric alternatives (11–13). GAM has been widely used in many time-series analyses, including those of data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS) (Moolgavkar S. Air pollution and daily mortality in three U.S. counties. Environ Health Perspect 2002;108:777–84). GAM extends traditional generalized linear models (GLM) (16) by replacing linear predictors of the form η = Σj βjxj with η = Σjfj(xj), where fj(xj) are unspecified nonparametric functions. Methods for estimating fj include smoothing splines or LOESS smoothers. GLMs with regression splines (to which we refer here as the fully parametric alternative of the GAM with nonparametric smoothers) commonly define fj to be regression splines, such as natural cubic splines or B-splines with a prespecified number of knots at known locations.

2.3.5. Empirical evaluation

The case fatality rate prediction was evaluated using four time-series models which were compared with the benchmarks. The benchmark is permitted to gauge its competitors’ impact (Kourentzes & Petropoulos, 2016). The SES model which allows for errors or trend elements is the most appropriate non-seasonal model for time series analysis (Haider et al., 2022). The execution of the time series models is examined and contrasted in this study to ensure robust prediction, coefficient of determination (R2), root mean square error (RMSE), and mean absolute error (MAE).

2.3.6. Statistical analysis

Spearman rank correlation coefficients were used to study meteorological variables and confirmed global daily Monkeypox cases. The evolution of MPX cases was examined using a time series model. We used the ARIMAX and negative binomial regression models to determine whether there is a link between MPX cases and deaths with meteorological variables using R software. The number of individuals with MPX infection is the dependent variable whereas meteorological variables are the independent variables in both models. The negative binomial regression (NBR) model was employed to further examine variations in MPX infections among nations. The Poisson-gamma mixed distribution is the foundation of a negative binomial regression model, which is helpful in forecasting count-based data. The adoption of this technique is due to the non-negative integer values of the reported monkeypox cases (the number of monkeypox infections) and the higher variance of the dependent variable compared to its mean. The association is validated using the coefficient, 95% confidence interval (CI), and the corresponding P-value.

Descriptive

|  |  |  |  |
| --- | --- | --- | --- |
| Variables | Mean ± SD | Minimum | Maximum |
| Temperature (°C) | 25.43 ± 4.18 | 15.39 | 31.86 |
| Dew/frost point temperature (°C) | 19.85 ± 6.30 | 6.69 | 26.83 |
| Relative humidity (%) | 74.76 ± 14.13 | 41.35 | 91.52 |
| Precipitation (mm/day) | 5.84 ± 5.82 | 0 | 28.39 |
| Surface pressure (kPa) | 100.68 ± 0.51 | 99.75 | 101.52 |
| Wind speed (m/s) | 1.94 ± 0.64 | 0.91 | 3.58 |
| Daily confirmed case (number of person) | 831.33 ± 3877.51 | 0 | 52636 |

According to meteorological factors in Table 1 the mean for dengue cases is 831.33, with a standard deviation (SD) of 3877.51, a minimum of 0, and a maximum of 52636 (Table 1).

Correlation plot



The Spearman rank correlation coefficients among meteorological variables and confirmed daily cases of Dengue suggest a significant but weak correlation between dengue cases and meteorological variables (Figure 1). The mean temperature and wind speed temperature exhibit an insignificant positive correlation with daily dengue cases (r = 0.095, p>0.05 and r = 0.013, p>0.05, respectively). How-ever, dew point, relative humidity and rainfall has a positive significant association with daily dengue cases (r = 166, p<0.05, r = 179, p<0.05 and r = 153. P<0.05), respectively. Surface pressure exhibit negative insignificant correlation with daily dengue cases (r = -0.104, p>0.05).

Time series plot of confirmed cases for different methods

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

We observed a constant trend between observed and predictive national dengue cases in the SES model, with R2, RMSE, and MAE of 21.14%, 3437.00, and 764.84, respectively (Table 2 and Figure 3). The prediction results from the time series models in Fig. 4 show confirmed and predicted dengue cases from January 2020 to November 2022. We observed a robust, increasing trend between observed and predictive dengue cases with R2, RMSE, and MAE of 35.74%, 3102.59, and 656.83 for the ARIMA model but 23.56%, 3383.94, and 656.83 for the SARIMA model and 37.16%, 3068.18, and 820.95 for the ARIMAX model (Table 2). The Prophet model performed weakly in terms of accuracy (i.e., the coefficient of determination is more significant along with lower or rates) compared to other models (i.e., R2 = 13.82%, RMSE = 3592.97, and MAE = 1510.02). However, The GA model performed most weakly (i.e., R2 = 10.20%, RMSE = 4012.51, and MAE = 1856.13). According to the forecast in all models except SES, the number of monkeypox cases is expected to increase considerably in next 30 subsequent days (Figure 3). In the M–K trend analysis, we identified an increasing trend of daily monkeypox cases (P<0.001 and tau = 0.268). Using Sen’s slope test, we find that the slope is 0.17 (95% CI: 0.09- 0.33) (Table 2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method & Period | *R2* | RMSE | | MAE |
| *Simple Exponential Smoothing* | | | | |
| SES | 21.14% | 3437.00 | | 764.84 |
|  |  |  | |  |
| *Auto-Regressive Integrated Moving Average* | | | | |
| ARIMA | 35.74% | 3102.59 | | 656.83 |
| *Seasonal Auto-Regressive Integrated Moving Average* | | | | |
|  | 23.56% | 3383.94 | | 736.54 |
| *Auto-Regressive Integrated Moving Average with explanatory variables* | | | | |
|  | 37.16% | 3068.18 | | 820.95 |
| *Automatic Forecasting time-series model* (Prophet Model) | | | | |
|  | 13.82% | 3592.97 | | 1510.02 |
| *Generalized Additive Model* | | | | |
|  | 10.20% | 4012.51 | | 1856.13 |
| *Mann-Kendell trend analysis* | | | | |
|  | Tau | | p-value | |
|  | 0.268 | | <0.001 | |
| *Sen’s slop test* | | | | |
|  | Sen’s Slope | | 95% CI | |
|  | 0.17 | | 0.09 to 0.33 | |

*RMSE: Root Mean Square Error; MAE: Mean Absolute Error*

In the ARIMAX and GA model, wind speed (-557.7 [95% CI: -1624.9 to 509.6] and -895.8 [-2312.1 to 520.4], respectively), dew point (-425.7 [-1649.3 to 797.9] and -765.1 [-2197.4 to 667.3], respectively) have an insignificant negative association with dengue cases, respectively. However, temperature and relative humidity are positively associated with dengue cases (363.7 [-793.0 to 1520.5] and 147.4 [-206.9 to 501.7], respectively) in the ARIMAX model, GA model also exhibit similar results, whereas rainfall are negatively associated with in ARIMAX model (-53.7 [-165.2 to 57.7] and positively associated 25.9 [-107.2 to 158.9], respectively) in the GA model (Table 3).

Factors associated with Dengue cases using the ARIMAX and GA model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Confirmed Cases | | | | | |
| Variables | ARIMAX Model | | | Generalized Additive Model | | |
|  | Coef. | 95%CI | P-value | Coef. | 95%CI | P-value |
| Wind Speed | -557.7 | -1624.9 to 509.6 | 0.306 | -895.8 | -2312.1 to 520.4 | 0.216 |
| Temperature | 363.7 | -793.0 to 1520.5 | 0.538 | 727.0 | -626.5 to 2080.5 | 0.293 |
| Dew point | -425.7 | -1649.3 to 797.9 | 0.495 | -765.1 | -2197.4 to 667.3 | 0.296 |
| Relative humidity | 147.4 | -206.9 to 501.7 | 0.415 | 237.3 | -170.1 to 644.7 | 0.255 |
| Rainfall | -53.7 | -165.2 to 57.7 | 0.345 | 25.9 | -107.2 to 158.9 | 0.703 |
| Surface pressure | -1659.4 | -4205.4 to 886.7 | 0.201 | -871.1 | -3948.9 to 2206.6 | 0.580 |